Ensemble-based parameter estimation in a coupled GCM using the Adaptive Spatial Average method

Y. Liu[*1], Z. Liu[*2,1], S. Zhang[3], X. Rong[4], R. Jacob[5], S. Wu[1], F. Lu[1]

1 Center for Climate Research and Dept. Atmospheric and Oceanic Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

2 Lab. of Ocean-Atmos. Studies, Peking University, Beijing 100871, China

3 GFDL/NOAA, Princeton University, Princeton, NJ 08542, USA

4 Chinese Academy of meteorological sciences, Beijing 100068, China

5 Mathematics and Computer Science Division, Argonne National Laboratory, IL 60439, USA

(to be submitted to J. climate)

*Corresponding authors:
Yun Liu, liu6@wisc.edu, Zhengyu Liu, zliu3@wisc.edu.

Abstract

Ensemble-based parameter estimation for a climate model is emerging as an important topic in climate research. For a complex system as a coupled ocean-atmosphere general circulation model, the sensitivity and response of a model variable to a model parameter could vary spatially and temporally. Here, we propose an adaptive spatial average (ASA) algorithm to increase the efficiency of parameter estimation. Refined from a previous spatial average method, the ASA uses the ensemble spread as the criterion for selecting "good" values from the spatially varying posterior estimated parameter values; the "good" values are then averaged to give the final global uniform posterior parameter. In comparison with existing methods, the ASA parameter estimation has a superior performance: faster convergence and enhanced signal-to-noise ratio.

1 Introduction

Parameter estimation using ensemble-based filter (Anderson, 2001) is emerging as a promising approach to optimize parameters in a complex model (Annan and Hargreaves, 2004; Hacker and Snyder, 2005; Annan et al., 2005 a & b; Ridgwall et al., 2007; Hacker and Snyder, 2005; Aksoy et al., 2006 a & b; Tong and Xue 2008 a & b; Nielsen-Gammon, 2010; Hu et al., 2010; Zhang el al, 2012; Zhang, 2011 a & b; Wu et al., 2012 a & b; Liu et al. 2014). In parameter estimation in a complex system, such as a coupled ocean-atmosphere general circulation model (CGCM), one common issue is sampling error accumulation when a large number of observations are used to update a single-value parameter sequentially (Aksoy et al, 2006a). To address this issue, Aksoy et al (2006a) proposed a spatial updating technique that transforms a single-value parameter into a two–dimensional field and updates the field spatially, so that localization in filtering can limits the observational error accumulation. The final model parameter after each analysis has been derived in two methods. In the first method, the globally uniform parameter value is recovered using a spatial average of the entire spatially varying parameter field (demoted as SA, Aksoy et al, 2006a & b). In the second method, the spatially varying parameters are allowed to vary spatially after each analysis, in the so-called Geographically-dependent Parameter Optimization (denoted as GPO, see Wu et al., 2012a).

Here, our objective is the recovery of the spatially uniform parameter value. We propose an average method called adaptive spatial average method (ASA). The ASA is refined from the SA method to increase the efficiency of parameter estimation. The ASA uses the ensemble spread as the criterion for selecting “good” parameter values from the spatially varying parameter estimation; these “good” values are then averaged to give the

final posterior parameter. Liu et al. (2014) has recently shown some examples of successful ASA estimation in a CGCM. In this paper, we will examine in detail the ASA methodology for parameter estimation in a CGCM using ensemble-based filter. The e-folding solar penetration depth (SPD) is used as the major parameter for estimation in this study. We will show that, compared with the SA method and the GPO method, our proposed ASA produces a faster convergence rate for parameter estimation. The paper is organized as follows. Section 2 briefly describes the parameter estimation scheme and the CGCM used in this study. Section 3 shows the model sensitivity to the parameter (SPD). Section 4 discusses the ASA method. The ASA method is compared with GOP method and SA method in section 5. A summary and further discussion are given in section 6.

2 Model and Method

(a) Fast Ocean Atmosphere Model (FOAM)

Our model, the Fast Ocean Atmosphere Model **(**FOAM, Jacob, 1997) is a CGCM with an atmospheric component of a R15 (7.5^o longitude, 4^o latitude and 18 layers). The ocean component is a z-coordinate model with a resolution of 2.8^o longitude, 1.4^o latitude and 24 layers. Without flux adjustment, the fully coupled model has been run for over 6000 years with no apparent drift in tropical climate (Liu et al., 2007a). In spite of its low resolution, FOAM has a reasonable tropical climatology (Liu et al., 2003), ENSO variability (Liu et al., 2000), and Pacific decadal variability (Wu et al., 2003, Liu et al., 2007b).

(b) Data Assimilation Scheme

We will use a particular EnKF scheme, the Ensemble adjustment filter (EAKF, Anderson, 2001, 2003) in this study. Model parameters will be estimated simultaneously

with the state variables by augmenting state variables with model parameters (Banks, 1992a, b; Anderson, 2001).

The e-folding solar penetration depth (SPD) is used as the major testing parameter for estimation. Solar attenuation in the ocean is a function of the amount of biomass in the upper layers of the ocean (Smith and Baker, 1978; Ohlmann et al., 2000). Previous studies suggest that solar penetration can have a significant impact on the surface climate in a climate model (Schneider and Zhu, 1998; Nakamoto et al., 2001; Murtugudde et al., 2002; Ballabrera-Poy et al., 2007; Anderson et al. 2007). In particular, some modeling studies found that a deeper solar attenuation leads to warming in the tropical Pacific annual mean SST, which may then reduce the cold bias in the equatorial Pacific in a coupled ocean-atmosphere model (Murtugudde et al., 2002; Ballabrera-Poy et al., 2007; Anderson et al. 2007).

Following Murtugudde et al. (2002), the downward solar radiation $I(z)$, at depth of z in FOAM is calculated as

$$I(z) = I(0)\gamma e^{\left(-\frac{z}{h}\right)} \qquad (1)$$

where $I(0)$ is the total incident solar radiation at sea surface and γ=0.47 (Frouin et al., 1989) represents the fraction of total solar radiation in the photosynthetically available radiation band (wavelengths from 380 to 700nm). The remaining fraction of solar radiance is fully absorbed in the top model layer of 20 meters. The h is the e-folding depth of the solar penetration depth (SPD), which will be estimated in our experiments. In the real world, the SPD can be treated as a state variable, too, because it can be calibrated using the remote sensing observation of ocean color. Here, however, it is

129 treated as a model parameter that will be estimated using conventional observation of sea
130 surface temperature (SST) and salinity (SSS).

131 In this paper, we assume the "truth" SPD has a globally uniform value of 17-m,
132 and the truth simulation is performed with this SPD. The first guess of SPD is assumed
133 20-m with an uncertainty of 3-m (standard deviation). The observation for the
134 assimilations are the monthly mean SST and SSS, which are generated by adding a
135 Gaussian white noise to the corresponding "truth" states at each grid point. The
136 observational error scales (standard deviation) are 1^{o}K for SST and 1psu for SSS. An
137 ensemble size of 30 is used in all of our experiments. A 30-year simulation from the
138 control truth run is used for the initialization of the ensemble, with the restart file of
139 January 1^{st} of each year used as the initial condition for each ensemble member. For state
140 variable, the upper 8 layers of ocean temperature and salinity (0~235m) are updated by
141 the observations. The Gaspari and Cohn (1999) covariance localization is used with the
142 influence radius of 3 grid points horizontally for both state variables and the parameter
143 SPD. To extract signal-dominant state-parameter covariance, the enhancive parameter
144 correction is applied (DAEPC, Zhang et al., 2012). Before the parameter estimation is
145 activated, the data assimilation is performed in a "spin-up" period of 2 years in which
146 only the state variables are estimated.

147 **3 Model sensitivity with respect to solar penetration depth**

148 We first investigate the model sensitivity to the parameter, solar penetration depth
149 (SPD). Two types of parameter sensitivities need to be considered when DAEPC is used
150 to improve the model climate. The first type is the sensitivity of the response of the
151 model climatology to the change of the parameter; this sensitivity shows if the final

model climate can be improved by tuning this specific parameter. The ocean surface climates of FOAM are significantly different between a deeper SPD (20-m) simulation and a shallow SPD (17-m) one, characterized by a warming of up to over 0.5^{o}K in the tropical ocean and a cooling of up to -0.5^{o}K in the subtropical ocean (see figure 1 in Liu et al. 2013).

The second type of sensitivity tests the model's sensitivity to parameter uncertainty (represented, say, by the ensemble spread of the parameter) in the observational space at the observational time interval; this sensitivity examines the possibility of reducing parameter uncertainty using the observations available. Furthermore, the model response to parameter uncertainty consists of linear and non-linear parts. Since the Kalman Filter framework is derived as the optimal analysis for a linear system, some features involving non-linear dependence may be regarded as noise for parameter estimation. Successful parameter estimation requires a signal-dominant state-parameter covariance, which is derived most favorably in a model whose state variables exhibit a strong linear dependence on model parameters (Aksoy et al., 2006 a, b).

An ensemble simulation starting from the same initial condition but using different values of the parameter SPD (i.e. an perturbed ensemble of parameters) demonstrates the second type of sensitivity (Fig. 1). (Here, the parameter ensemble is constructed as a Gaussian distribution with the mean of 20-m and the standard deviation of 3-m). Since we will use the observations of monthly SST for parameter estimation, we will examine the ensemble response of the first month SST. The ensemble spread of the first month SST (monthly mean) represents the response of the model SST to the

175 uncertainty of SPD in the observational space; the correlation coefficient between the
176 SPD ensemble and the first month SST quantifies the linear part of the response. Fig.1
177 shows an overwhelmingly negative correlation between SST and SPD, implying
178 predominantly a colder SST with a deeper SPD. This cooling is likely to be caused by the
179 direct effect of solar penetration. Physically, a deeper SPD allows more solar radiation to
180 penetrate below the surface layer, leaving less shortwave radiation heating the surface
181 layer, and therefore cause surface cooling. The direct effect of solar penetration is
182 dominant in the initial months in response to a sudden change of the SPD (Hokanson,
183 2006). One striking feature of the sensitivity is the strong variation with season and
184 location. The SST ensemble spread is large and exhibits negative correlations in the
185 summer hemisphere where the mixed layer is shallow and therefore the SST is more
186 sensitive to heat flux perturbations. Fig.1 is important for our parameter estimation,
187 because it indicates the key regions for parameter estimation. The regions with large
188 sensitivity and high correlation represent the regions of large linear model response to
189 SPD. These regions have high signal/noise ratio and therefore are the regions where the
190 observation of SST are most effective for parameter estimation. The rest of regions,
191 which account for more than half of the grid points at each analysis step, are unlikely to
192 provide significant information for parameter estimation.

193 **4 The Adaptive Spatial Average scheme (ASA)**

194 The sensitivity experiments in section 3 show that the model response to the
195 parameter SPD varies significantly in both space and time. We speculate that neither
196 GPO nor SA is most efficient for estimating the parameter. This follows that only the
197 regions with large model-to-parameter linear response can provide state-parameter

198 covariance with high signal/noise ratio for parameter estimation. Fig.3 implies that the
199 state-parameter covariance is insignificant over about half of the grid points at a time and
200 in about half of the year at a given grid point. Therefore, for the purpose of parameter
201 estimation, the estimations are not useful for more than half of the time at a given grid,
202 and the estimations are not useful for more than half of the grids in the basin for a given
203 observation time. Therefore, both SA and GPO are not the most efficient methods to
204 estimate the parameter SPD, as will be shown below.

205 Here we refine the SA method to the Adaptive Spatial Average (ASA) method, to
206 increase the efficiency of parameter estimation. In SA, the final spatially uniform
207 parameter is estimated as the average of all the spatially different posteriors, each derived
208 at a grid point using localization. The ASA is based on the idea that a parameter
209 estimation, which will be derived from an average of spatially different posteriors, should
210 be more accurate if it only includes average those posteriors of smaller uncertainties (i.e.
211 errors). For practical applications where the truth parameter, and therefore, the parameter
212 error, is unknown, we can consider the ensemble spread as a representation of the error,
213 as in traditional application of ensemble filtering to state variables (e.g. Evensen, 2007).
214 (We will return to this point later). Therefore, the ensemble spread can be considered as
215 the indicator of the quality of each posterior parameter values and a higher quality
216 posterior has a smaller ensemble spread. The ASA will only retain those high quality
217 values for the final averaging to derive the value for the spatially uniform parameter.
218 This average value of high quality values should have smaller error than the average
219 value of averaging all the values as in SA, which include the high quality as well as low
220 quality values. A preliminary theoretical analysis of this point is given in the appendix.

A posterior value is "good" if its ensemble spread is relatively small among all the posteriors estimated at all the grid points. In practice, we use a threshold of the spread ratio between the posterior and the prior to judge the quality of the posterior and a posterior with a spread ratio below the threshold is considered a "good" posterior to be included for the final spatial average. (It should be noted that the ensemble spread of the prior is spatially uniform over the globe. Therefore, this spread ratio of the posterior over prior does not affect the relative magnitude of the posterior.) The speed of the decrease of the parameter uncertainty depends greatly on the magnitude of the signal. Initially, the ASA can use a small ratio as the threshold because the initial parameter uncertainty is large and the response magnitude (signal) is large. The threshold will be increased during the simulation with the decrease of the parameter uncertainty. The ASA is applied every few EnKF analysis cycles to obtain sufficient numbers of "good" parameter posterior values. The ASA therefore differs from the SA of Aksoy et al (2006a), in which the spatial average is performed every EnKF analysis cycle and on all grid points. A conditional covariance inflation technique (CCI) as in Aksoy et al. (2006b) is also employed here on parameter ensemble after each ASA step to avoid the filter divergence for parameter estimation. The CCI inflates the parameter ensemble back to a predefined minimum value when necessary. The predefined minimum value is also the final uncertainty target for the estimated parameter.

5 Comparison of ASA with GPO and SA

We now compare ASA with SA and GPO schemes in FOAM. Two sets of experiments of parameter estimation are performed using observations of monthly SST

and SST at every grid point. The first set of experiments (EXP-1a and EXP-1b) use the GOP scheme and confirm that the parameter ensemble spread is a good index for the parameter uncertainty (Figs. 2, 3). The second sets of experiments (EXP-2a and EXP-2b, Figs. 4, 5) compare the parameter estimations between SA and ASA schemes. The details of experimental setting are shown in table 1.

(a) The assimilations with GPO scheme

Both EXP-1a and EXP-1b use the GPO scheme but with different observations. EXP-1b uses regular observations that consist of the "truth" plus noise. EXP-1a, called perfect observation experiment, uses the "truth" from control as the observations but nevertheless treats it as having the same uncertainty scale as in EXP-1b. For these two GPO experiments, neither EXP-1a nor EXP-1b is able to produce good parameter estimation, if only the monthly SST and SSS data are assimilated. Therefore, we are forced to also assimilate daily atmosphere wind (U, V) and temperature (T) with the error scale of 1 m/s and 1K, respectively; the observational error scales for SST and SSS are also forced to be reduced from 1^{o}K and 1 psu to 0.5^{o}K and 0.5 psu, respectively. The initial SPD error is also reduced from 3-m to 1-m.

As speculated, the spatial pattern of the RMSE of SPD in EXP-1a is very consistent with the ensemble spread after 20 years of simulation (Figs. 2a, b). There are some regions of low uncertainty of SPD in different ocean basins. A further study shows that the low uncertainty in the mid-latitude North Pacific and North Atlantic are related to the large model sensitivity to SPD during the boreal summer (Fig1b) and fall (Fig1c); the low uncertainty in the eastern South Pacific, western equatorial Pacific, South Atlantic and southern Indian Ocean are partly related to the large sensitivity of the model SST to

267 SPD in the austral fall (Fig.1a) and summer (Fig.1d). The high positive correlation
268 between the parameter uncertainty and its ensemble spread can be seen more clearly in
269 the scatter plot, for example, at the simulation year of 40 (Fig3a). The RMSE of SPD
270 estimation and its ensemble spread show a strong positive linear correlation with only
271 modest spread residual. The estimate values are closer to the truth when the ensemble
272 spread is small, except for the case of very small ensemble spread (<~ 0.3 in Fig.3a). The
273 positive correlation between the posterior error and ensemble spread supports our
274 speculation before that the ensemble spread can be used to represent the estimation error
275 or uncertainty. Furthermore, it is clear that a spatial average will decrease the parameter
276 error because the average reduces the part of parameter uncertainty that is spatially
277 independent (see eqn. (A4 in the appendix). The error of SPD can be further reduced by
278 using only the posterior values with smaller ensemble spread for average (Fig3b), as
279 hypothesized for the ASA. The error of SPD is reduced to 0.40-m when the posterior
280 values of SPD over all the global grid points are averaged in EXP-1a (after 40 years of
281 assimilation), compared with the global mean RMSE of SPD of 0.6-m (first RMSE and
282 then global average); this error is decreased to 0.2-m and 0.1-m when the top 50% and
283 20% of grid points of smallest ensemble spread are averaged, respectively. When the
284 ensemble spread is at its smallest values, the estimated values suffer from an overshoot,
285 i.e. the parameter error becomes negative. This phenomenon also occurs in Liu et al.
286 (2013) when the similar observation coverage is applied, i.e. U, V and T for the
287 atmosphere and SST and SSS for the ocean. The reason for the overshoot will be
288 discussed in a future study.

289 The positive correlation between the parameter uncertainty and parameter
290 RMSE, however, is disrupted significantly when the regular observation ("truth" plus
291 noise) is used as in EXP-1b. Now, the spatial pattern of the parameter ensemble spread
292 (Fig.2d) remains similar to that in EXP-1a (Fig.2b), but the pattern of the SPD
293 uncertainty (Fig.2c) become very noisy. This occurs because the parameter updating
294 using EnKF also introduces observational errors into the SPD posterior, which is
295 equivalent to adding random noise onto the parameter posterior of EXP-1a. This noise
296 leads to a decrease of the consistence between the SPD uncertainty and its ensemble
297 spread. The distortion on the correlation is seen clearly in the scatter plot Fig3c, where
298 the error value of SPD and its ensemble spread of EXP-1b shows a very weak linear
299 relationship with a much-enhanced residual variance. Nevertheless, this correlation is still
300 significant at the 99% level. Furthermore, since the uncertainty associated with the
301 observation errors is spatially independent, it can be reduced dramatically using a spatial
302 average. Indeed, the averaging values of SPD are very similar for EXP-1a and EXP-1b
303 (Figs.3b vs. d), although the estimated values of SPD are much more noisy in EXP-1b
304 than in EXP-1a.

305 Overall, the consistency between the parameter uncertainty and its ensemble
306 spread indicates that the parameter ensemble spread can be used as a good index for the
307 uncertainty of the parameter value and therefore can be used as the criteria for selecting
308 "good" posteriors for averaging. A spatial average of those "good" posteriors tends to
309 give a better final estimation.

310 **(b) Comparison between SA and ASA**

311 As discussed regarding EXP-1a, 1b, and in the appendix, the uncertainty of the
312 parameter posterior can be reduced using spatial average. The ASA and SA are applied in
313 EXP-2a and EXP-2b, respectively. A predefined minimum ensemble spread value of 0.3-
314 m for the CCI is applied in the EXP-2 (s). Unlike the GPO experiments above, now, the
315 error of SPD is reduced dramatically in both EXP-2a and EXP-2b even only with
316 monthly mean SST and SSS observations (Fig. 4a), implying an increased robustness of
317 parameter estimation using spatial average.

318 Based on the ensemble sensitivity shown in Fig.1, we apply the ASA every 6
319 analysis cycles (6 months) in EXP-2a with an initial threshold of 0.68. To prevent the
320 degeneration case of too few "good" values, the threshold increases by 0.1 until it reaches
321 0.98 whenever the total number of "good" values is smaller than a given number, here set
322 as 400. The ASA picks different grids at different times for averaging. The number of
323 grid points of "good" values also varies temporally in the range of 400~4000, which is
324 around 2~40% of total ocean grids (Fig 4b). The ensemble spread of SPD initially
325 decreases much faster than its real uncertainty (Fig4a), reaching the minimum parameter
326 ensemble spread of 0.3-m in 5 simulation years. Although this ensemble spread (0.3) is
327 smaller than the real error in years of 5-20, the SPD continues to converge to its "truth".
328 The SPD error in EXP-2a is decreased from 3-m to 0.3-m (the estimating goal) in 20
329 years (Fig 4a).

330 During the assimilation cycle, the ensemble spread still remains positively
331 correlated with the estimation errors among different points, albeit with a substantial
332 spread (as discussed for Exp.1b in Fig.3b). This can be seen in the two examples of
333 scatter plots of SPD after the first and fifth spatial updating cycles in Fig.5a and 5b,

334 respectively. The ASA produces a good SPD estimation by averaging only a moderate
335 number of "good" values (200-2000) once the threshold (the uncertainty ratios between
336 the posterior and prior) is selected appropriately. This can be seen in Figs.5c and Fig.5d,
337 which shows the number of "good" values and the average of these "good" values
338 respectively, as functions of the threshold in ASA for the first 5 assimilation cycles. For
339 example, for the 1st cycle, the average of SPD is 18.5 with the threshold of 0.8 and the
340 number of "good" value of ~400; and the average of value is 17.6 with the threshold of
341 0.65 and the number of "good" value of ~1000. If the threshold is too small, too few
342 values are defined as "good" values. This will lead to a too small sample size and large
343 sampling error, such that ASA no longer produces good results (Fig. 5b & d).
344 The final estimation also depends on the minimum ensemble spread specified in
345 CCI. The error of the estimated SPD seems to saturate at the equilibrium level of ~0.2-m
346 error in ~30 years in EXP-2a if the minimum parameter ensemble spread remains at 0.3-
347 m. This minimum ensemble spread can be decreased afterwards to yield more accurate
348 estimation. The ASA estimation is repeated from year 31 to year 47 but now with the
349 minimum parameter ensemble spread reduced from 0.3-m to 0.2-m; now the SPD error
350 further decreases from 0.2-m to ~0.1-m (Fig. 4a,green lines).). In this case, a reduced
351 minimum ensemble spread further improves the final convergence of the parameter
352 estimation.
353 In comparison with the ASA (in EXP-2a), the spatial average using all the grid
354 points in SA (EXP-2b) shows a considerably slower convergence in the SPD estimation,
355 with the SPD error barely reaching 0.3-m after 47-years of assimilation (red lines, Fig 4a).
356 Similar to the ASA, the ensemble spread of SPD in SA also decreases much faster than

its real error scale. The CCI with the minimum parameter ensemble spread of 0.3 prevents the filter divergence of the parameter estimation. In the mean time, the evolution of estimation SPD in SA is more stable than in ASA because more grids and in turn a bigger sample size in the former than the latter [also see eqn. (A14) in appendix]. Overall, ASA demonstrates a faster convergence rate than SA for SPD estimation because the former uses only "good" values for averaging.

5. Summary and Discussions

Refining the Spatial Average scheme (SA), we proposed the Adaptive Spatial Average scheme (ASA) to improve the efficiency of the parameter estimation in a complex system, such as a CGCM. The ASA is explored in the twin experiment framework in FOAM, where the biased parameter (SPD) is the only model error source. The e-folding scale of the solar penetrating depth is used as the biased parameter for estimation. Sensitivity experiments show that the response of the FOAM to the parameter uncertainty varies spatially and temporally. The ASA is demonstrated to increases the efficiency of parameter estimation significantly over previous assimilation techniques such as the SA (Aksoy et al., 2006a) and geographic dependent parameter optimization (GPO) (Wu et al, 2012a).

The ASA uses the posterior ensemble spread as the criterion to select the "good" values from the spatial updating posterior parameter values and only use the "good" values for the averaging to yield the globally uniform posterior. In comparison with the SA scheme, the ASA produces a faster convergence for parameter estimation. The faster convergence of ASA than SA is robust in other settings, as seen in two additional pairs of experiments the same as EXP-2a & b, except for the observational interval of 10 days

380 (EXP-3a & b) and 1 day (EXP-4a & b), respectively (Table 1). When the observational
381 interval is shortened, the model response to the parameter uncertainty becomes more
382 linear. However, the response amplitude still varies spatially and temporally (not shown).
383 Therefore, ASA is still more suitable than SA. Similar to EXP-2, both EXP-3 and EXP-4
384 show faster decreases of the SPD ensemble spread than its real uncertainty in the initial
385 stage. The convergence time is also shortened for a shorter observational interval. In
386 ASA, the SPD errors reach the objective uncertainty (0.3-m) in ~10 years (EXP-3a,
387 Fig.6a) and ~5 years (EXP-4a, Fig.6b) of simulations, for the observational interval of 10
388 and 1 days, respectively, while, in SA, they take ~30 years (EXP-3b, Fig.6a) and ~10
389 years (EXP-4b, Fig.6b). It is noted that the estimated SPD in EXP-4 (Fig.6b) is less
390 stable than that in EXP-2 or EXP-3 (Fig.3a, Fig.6a). The observational interval in EXP-4
391 is only 1 day, while the decorrelation time scale of SST is a few months. This results in
392 the accumulation of sampling error because the model SST does not have the time to
393 respond before another observation is added. The accumulation of sampling error causes
394 poor parameter estimation compared to the other experiments. Furthermore, the
395 instability of the estimated parameter in Fig.6b could become worse as the total
396 assimilation time increases. We could increase the assimilation time interval for
397 parameter estimation to reduce the instability of parameter estimation.

398 The ASA is designed to deal with the spatially and temporally varying feature of
399 model response to parameter in CGCM. As pointed out by one reviewer, for SPD, SST
400 shows little sensitivity to the parameter perturbation in about half of the world ocean
401 (Figs.1a-d). One may speculate that our experiments for the estimation of SPD are too
402 peculiar. The SA is inferior to ASA because the posteriors in these regions of little

403 sensitivity are subject to too large a noise (with little response signal) and therefore
404 contaminate the SA estimation seriously. To clarify this, it will be desirable to test the
405 estimation for a parameter that has a more spatially uniform response sensitivity.
406 Therefore, we repeated the estimation for two other parameters m_d and m_q (also see Liu
407 et al., 2014). The m_d and m_q are artificial multipliers to the momentum and latent heat
408 fluxes between the ocean and atmosphere, respectively, with 1 as the default truth model
409 value. The model SST sensitivity to either parameter is more uniform than for SPD (not
410 shown). Our experiments EXP-5a and b and EXP-6a and b use the same experimental
411 setting as EXP-2a and b except for estimating the imperfect parameter m_d and m_q,
412 respectively (Table1, Fig.7). Both EXP-5a and EXP-6a show faster decreases of the
413 parameter errors than EXP-5b and EXP-6b. The m_d reaches the objective uncertainty of
414 0.04 (set by the minimum ensemble spread specified in CCI) in ~10 years with ASA but
415 in more than 30 years of assimilation with SA (Fig.7a). Similarly, the m_q reaches the
416 objective uncertainty of 0.04 in ~25 years with ASA but in more than 40 years of
417 assimilation with SA (Fig.7b). Therefore, the improvement of ASA over SA is valid for
418 more general cases than the SPD.

419 The ASA has also been shown successful for the estimation of multiple
420 parameters (Liu et al., 2014). Therefore, we believe that the ASA method is well suited
421 for the estimation of those parameters with a globally uniform feature in CGCM. The
422 estimation of a spatially varying parameter in CGCM, however, remains to be further
423 studied.

424 Much further work remains. All of our experiments of parameter estimation in
425 this study were implemented in a twin experiment framework, where the sampling error

426 is one of the major error sources for parameter estimation. The parameter estimation
427 using the real observational data will be much more complex than that. Aside from the
428 parameter uncertainties, the model bias can be generated in a CGCM due to model
429 structural errors, such as the imperfect dynamical framework and the incomplete
430 understanding for physical processes. It remains a great challenge to identify the sources
431 of the model bias from the candidates of the model structural deficiencies, as well as the
432 large number of model parameters. Hu et al (2010), in their real-data parameter
433 estimation study, pointed out that the parameter estimation using real observations might
434 produce the right answer for the wrong reasons. Furthermore, the uncertainty generated
435 by the model structural errors cannot be included in a single model ensemble forecast.
436 Therefore, the background uncertainty estimated from the ensemble perturbations usually
437 suffers a negative deficiency when we apply parameter estimation using real
438 observations. A negatively biased background uncertainty could cause poor filter
439 performance or even filter divergence, and therefore cause parameter estimation failure.
440 One has to tune the inflation factor to compromise the uncertainty deficiency using a
441 state-of-the-art inflation schemes, such as the covariance inflation/relaxation (Zhang et
442 al., 2004), the additive inflation (Hamill and Whitaker, 2005), or the adaptive covariance
443 inflation (Anderson, 2007, 2009).
444
445 **Acknowledgment**
446 We gratefully appreciate Ms. M. Kirchmeier for her help in editing the
447 manuscript. We would also like to thank two anonymous reviews for their comments on
448 an earlier version of the manuscript. We gratefully acknowledge the computing resources

449 provided on "Fusion," a 320-node computing cluster operated by the Laboratory
450 Computing Resource Center at Argonne National Laboratory. This research is sponsored
451 by NSF and Chinese MOST 2012CB955200. This work was supported in part by the
452 U.S. Department of Energy, Office of Science, under Contract DE-AC02-06CH11357.
453 This paper is CCR contribution #xxxx
454

455 **Appendix: Preliminary Theoretical Consideration for ASA**

456 Here, we will discuss the SA and ASA from a more quantitative perspective.
457 When we implement the spatial updating in ensemble-based parameter estimation, we
458 obtain a spatially varying parameter posterior field. The posterior errors at different
459 locations are correlated because the parameter priors are identical for the entire field. To
460 quantify the effect of spatial averaging, we can separate the posterior errors into two
461 independent components: one linearly dependent on the parameter prior error and the
462 other uncorrelated with the first one.

463 In EnKF, the covariance(s) between the parameter and the model forecasts in
464 observational space are used directly to update parameter in exactly the same manner as
465 for the state variables. When we use a forecast x^f and an observation x^o to update a
466 parameter β, the $(\sigma_\beta^2)_a$ of a parameter posterior can be written as

467 $$(\sigma_\beta^2)_a = \sigma_\beta^2(1-\theta) \quad \text{(A1)}$$

468 where $\theta = \frac{\rho^2\sigma_x^2}{(\sigma_x^2+R)}$ with $0 \le \theta < 1$. Here the σ_x^2, R are the error scales (variances) of x^f
469 and x^o, respectively; ρ is the correlation coefficient between forecast x^f and parameter
470 prior. The uncertainty of parameter posterior decreases with the increase of θ. The ratio
471 between parameter posterior uncertainty and prior uncertainty

$$\frac{(\sigma_\beta^2)_a}{\sigma_\beta^2} \equiv r = 1 - \theta.$$

In EnKF, $(\sigma_\beta^2)_a$ and σ_β^2 are represented by the variance of parameter posterior and prior ensemble, respectively. So the r is the ratio between the posterior and the prior ensemble spread. For a spatial updating, different location has different r. The ASA uses the r as index to select the "good" values from a posterior field.

The parameter posterior error of ε_β^a originates from different sources: x^f, x^o and β^f, and can be written into two parts based on the correlation relationships among the error sources

$$\varepsilon_\beta^a = \sigma_\beta N_\beta^f (1-\theta) + \sigma_\beta \sqrt{\theta - \theta^2} N_x^b \quad \text{(A2)}$$

where N_β^f and N_x^b are independent white noise with the scale of 1. The two terms on the right hand side of (A2) represent two independent components of the total uncertainty (error) of a posterior value for any given σ_β and θ. The 1st term linearly depends on the error of parameter prior of $(\sigma_\beta N_\beta^f)$, while the 2nd term is uncorrelated with the error of parameter prior. The 2nd term is produced by the errors from observations, initial conditions and the nonlinear part of model response to the parameter prior. The 1st term is dominant when θ is close to 0 and the uncertainty of posterior is close to the uncertainty of parameter prior. The second term become primary when θ is close to 1 and the uncertainty of posterior is much smaller than the uncertainty of parameter prior (Fig. 8).

For a spatial updating, we can rewrite (A2) into a spatially varying field

$$\varepsilon_{\beta,i}^a = \sigma_\beta N_\beta^f (1-\theta_i) + \sigma_\beta \sqrt{\theta_i - \theta_i^2} N_{x,i}^b \quad \text{(A3)}$$

where $i = 1, 2 \ldots, N$ indicate the locations. The first term on the right hand side is all linearly dependent among different locations, while the second term on the right hand side can be regard as independent among different locations when the posterior values are widely distributed over a large domain. For a spatial average, the two terms have opposite changes. Averaging the β_i^a to obtain a single-value parameter, the posterior error is

$$\overline{\varepsilon_{\beta,1}^a} = \frac{\sigma_\beta}{M}\sum_i N_\beta^f(1-\theta_i) + \frac{\sigma_\beta}{M}\sum_i\left(\sqrt{\theta_i - \theta_i{}^2}N_{x,i}^b\right) \quad \text{(A4)}$$

We now discuss the two terms on the right hand side of (A4) one by one, regarding the difference between SA and ASA. The first term is linearly dependent on the parameter prior error (N_β^f), therefore its scale mainly affect by the distribution of θ_i but not the averaging sample size of M. The first term can be discussed conveniently by assuming a uniform distribution $[\theta_{min} \quad \theta_{max}]$ for θ_i. The SA scheme (Aksoy et al 2006a) averages all posterior values over the entire domain. This term becomes $\sigma_\beta(1-\frac{\theta_{min}+\theta_{max}}{2})N_\beta^f$. The ASA sets a threshold θ_{th} ($\theta_{min} \le \theta_{th} \le \theta_{max}$) to remove the values with $\theta_i < \theta_{th}$ from the average pool such that this term becomes $\sigma_\beta(1-\frac{\theta_{th}+\theta_{max}}{2})N_\beta^f$, which is smaller than that using the SA scheme when the difference between θ_{min} and θ_{max} is large and θ_{th} is significantly greater than θ_{min}. When $\theta_{th} = \theta_{min}$, the ASA recovers to the SA. When $\theta_{th} = \theta_{max}$, the ASA just picks the posterior value with the "best" posterior, i.e. the minimum analysis error.

The second term on the right hand side of (A4) decreases with the increase of the average sample size of M because the $N_{x,i}^b$ are independent among different sites. Therefore the second term in ASA is larger than that in SA because ASA uses a smaller M than SA. However, when the number of average values (M) is sufficiently large, the

second term for both SA and ASA is smaller than the first term (unless the $\theta_i(s)$ are all close to 1), and therefore has limited impact on the total error. When the $\theta_i(s)$ are all close to 1, the first term is trivial comparing with the second term before average (see (A2) and Fig.8); but this rarely happens for parameter estimation with EnKF in a complex system like CGCM, because it would require $\rho^2 \approx 1, \sigma_x^2 \gg R$. The θ_{min} is usually close to 0, especially when the parameter is nearly converging.

The ASA can reduce the error related to the parameter prior error in spite of a reduced the averaging sample size, because "good" posteriors are used which have sufficiently large θ_i. The ASA produces better analysis of β than SA when the θ_{th} (θ_{max}) is significantly larger than the θ_{min} when they average the same posterior field. In summary, the SA reduces the errors related to the observations and forecasts. These errors are uncorrelated between different locations. The ASA scheme enhances the signal during the averaging by filtering out the region with weak signal or no signal. Therefore the ASA can produce a faster convergence than the SA (see Fig.4a, Fig.6 and Fig.7).

Reference

Aksoy, A., F. Zhang and J. W. Nielsen-Gammon, 2006a: Ensemble-based simultaneous state and parameter estimation with MM5. Geophys. Res. Lett., 33, L12801, doi:10.1029/2006GL026186.

Aksoy, A., F. Zhang and J. W. Nielsen-Gammon, 2006b: Ensemble-based simultaneous state and parameter estimation in a Two-Dimensional Sea-Breeze Model. Mon. Wea. Rev., 134, 2951-2970.

Anderson, J. L., 2003: A local least squares framework for ensemble filtering. Mon. Wea. Rev., 131, 634-642.

537 Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. Mon.
538 Wea. Rev., 129, 2884-2903.
539 Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithm for
540 ensemble filters. Tellus, 59A, 210-224.
541 Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for
542 ensemble filters. Tellus, 61A, 72-83.
543 Anderson, W. G., A. Gnanadesikan, R. Hallberg, J. Dunne, and B. L. Samuels, 2007:
544 Impact of ocean color on the maintenance of the Pacific cold tongue. Geophys. Res.
545 Lett., 34, L11609, doi: 10.1029/2007GL030100.
546 Annan, J. D. and J. C. Hargreaves, 2004: Efficient parameter estimation for a highly
547 chaotic system. Tellus, 56A, 520-526.
548 Annan, J. D., J. C. Hargreaves, N. R. Edwards and R. Marsh, 2005a: Parameter
549 estimation in an intermediate complexity Earth System Model using an ensemble
550 Kalman filter. Ocean Modelling, 8, 135-154.
551 Annan, J.D., D.J. Lunt, J.C. Hargreaves, and P. J. Valdes, 2005b: Parameter estimation in
552 an atmospheric GCM using the ensemble Kalman filter. Nonlinear Processes
553 Geophys. 12, 363-371.
554 Banks, H.T., 1992a: Control and estimation in distributed parameter systems. In: H.T.
555 Banks, Editor, Frontiers in Applied Mathematics vol. 11, SIAM, Philadelphia,
556 pp227.
557 Banks, H.T., 1992b: Computational issues in parameter estimation and feedback control
558 problems for partial differential equation systems. Physica D 60, 226-238.

559 Ballabrera-Poy, J., R. Murtugudde, R.H. Zhang, and A.J. Busalacchi 2007: Coupled
560 ocean-atmosphere response to seasonal modulation of ocean color: Impact on
561 interannual climate experiments in the tropical Pacific. J. Clim., 20, 353-374.
562 Evensen, G., 2007: Data Assimilation: The Ensemble Kalman Filter, Springe Press,
563 pp187.
564 Frouin, R., D. W. Lingner, C. Gautier, K. S. Baker, and R. C. Smith 1989: A simple
565 analytical formula to compute clear sky total photosynthetically available olar
566 irradiance at the ocean surface. J. Geophys. Res., 94, 9731-9742.
567 Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three
568 dimensions. Quart. J. Roy. Meteor. Soc.,125, 723–757.
569 Hamill, T. M., and J. S. Whitaker, 2005: Accounting for the error due to unresolved
570 scales in ensemble data assimilation: A comparison of different approaches. Mon.
571 Wea. Rev., 133, 3132–3147.
572 Hacker, J.P. and C. Snyder, 2005: Ensemble Kalman filter assimilation of fixed screen-
573 height observations in a parameterized PBL. Mon. Wea. Rev., 133, 3260-3275
574 Hokanson, Erin P., 2006: The effects of solar penetration on a coupled general circulation
575 model, masters theses, UW-madison, department of atmospheric oceanic sciences.
576 Hu, X.-M. , F. Zhang, and J. W. Nielsen-Gammon, 2010: Ensemble-based simultaneous
577 state and parameter estimation for treatment of mesoscale model error: A real-data
578 study, Geophys. Res. Lett., 37, L08802, doi:10.1029/2010GL043017.
579 Jacob, R., 1997: Low frequency variability in a simulated atmosphere-ocean system.
580 Ph.D thesis, University of Wisconsin-Madison, 155pp

581 Kalman, R., 1960: A new approach to linear filtering and prediction 644 problems. Trans.
582 ASME, Ser. D, J. Basic Eng., 82, 35-45.
583 Kalman, R. and R. Bucy, 1961: New results in linear filtering and prediction theory.
584 Trans. ASME, Ser. D, J. Basic Eng., 83, 95-109.
585 Kondrashov, D., C. Sun and M. Ghil, 2008: Data assimilation for a coupled ocean
586 atmosphere model, Part II: Parameter estimation. Mon. Wea. Rev., 136, 5062-5076
587 Liu, Y. Z. Liu, S. Zhang, F. Lu, X. Rong, R. Jacob, S. Wu, 2013: Ensemble-based
588 Parameter Estimation in a Coupled General Circulation Model. Submitted to J.
589 Climate
590 Liu, Z., J. Kutzbach and L. Wu, 2000: Modeling climatic shift of El Nino variability in
591 the Holocene. Geophys. Res. Lett., 27, 2265-2268.
592 Liu, Z., B. Otto-Bliesner, J. Kutzbach, L. Li, C. Shields, 2003: Coupled climate
593 simulations of the evolution of global monsoons in the Holocene. J. Clim., 16, 2472-
594 2490
595 Liu, Z., Y. Wang, R. Gallimore, F. Gasse, T. Johnson, P. deMenocal, J. Adkins, M.
596 Notaro, I.C. Prentice, J. Kutzbach, R. Jacob, P. Behling, L. Wang and E. Ong, 2007a:
597 Simulating the transient evolution and abrupt change of Northern Africa
598 atmosphere–ocean–terrestrial ecosystem in the Holocene. Quat. Sci. Rev.,26, July
599 2007, 1818-183, doi:10.1016/j.quascirev.2007.03.002
600 Liu, Z., Y. Liu, L. Wu and R. Jacob, 2007b: Seasonal and Long-Term Atmospheric
601 Responses to Reemerging North Pacific Ocean Variability: A Combined Dynamical
602 and Statistical Assessment. J. Climate, 20, 955–980.

603 Murtugudde, R., J. Beauchamp, and A. Busalaccki 2002:, Effects of penetrative radiation
604 on the upper tropical ocean circulation. J. Clim., 15, 470-486.
605 Nakamoto, S., S. Prasanna Kumar, J. M, Oberhuber, J. Ikshizaka, K. Muneyama, and R.
606 Frouin 2001: Response of the equatorial Pacific to chlorophyll pigment in a mixed
607 layer isopycnal ocean general circulation model. Geophys. Res. Lett, 28, 2021-2024.
608 Nielsen-Gammon, J. W., X.-M. Hu, F. Zhang, and J. E. Pleim, 2010: Evaluation of
609 Planetary Boundary Layer Scheme Sensitivities for the Purpose of Parameter
610 Estimation, Mon. Wea. Rev., 138, 3400-3417.
611 Ohlmann, J. C., D. A. Siegel, and C. D. Mobley 2000: Ocean radiant heating. Part I;
612 Optical influences. J. Phys. Oceanogr., 30, 1833-1848.
613 Ridgwell, A., J. C. Hargreaves, N.R. Edwards, J.D. Annan, T. M. Lenton, R. Marsh, A.
614 Yool and A. Watson, 2007: Marine geochemical data assimilation in an efficient
615 Earth System Model of global biogeochemical cycling. Biogeosciences, 4, 87-104
616 Schneider, E., and Z. Zhu 1998: Sensitivity of the simulated annual cycle of sea surface
617 temperature in the equatorial Pacific to sunlight penetration. J. Clim., 11, 1932-1950.
618 Smith, R. C. and K. S. Baker 1978: The bio-optical state of ocean waters and remote
619 sensing. Limnol. Oceanogr., 23, 247-259.
620 Tong, M. and M. Xue, 2008a: Simultaneous estimation of microphysical parameters and
621 atmospheric state with simulated Radar data and ensemble square root Kalman filter.
622 Part I: Sensitivity analysis and parameter identifiability. Mon. Wea. Rev., 136, 1630-
623 1648.

624 Tong, M. and M. Xue, 2008b: Simultaneous estimation of microphysical parameters and
625 atmospheric state with simulated Radar data and ensemble square root Kalman filter.
626 Part II: Parameter estimation experiments. Mon. Wea. Rev., 136, 1649-1668.
627 Wu. L, Z. Liu, R. Gallimore, R. Jacob, D. Lee, and Y. Zhong, 2003: A coupled modeling
628 study of Pacific decadal variability: The Tropical Mode and The North Pacific Mode.
629 J. Climate, 16, 1101-1120
630 Wu, X., S. Zhang, Z. Liu, A. Rosati, T. Delworth and Y. Liu, 2012a: Impact of
631 Geographic Dependent Parameter Optimization on Climate Estimation and
632 Prediction: Simulation with an Intermediate Coupled Model. Mon. Wea. Rev., **140**,
633 3956–3971.
634 Wu, X., S. Zhang, Z. Liu, A. Rosati, and T. Delworth, 2012b: A study of impact of the
635 geographic dependence of observing system on parameter estimation with an
636 intermediate coupled model. Climate dynamics. Doi:10.1007/s00382-012-1385-1.
637 Zhang, F., C. Snyder, and J. Sun, 2004: Impact of initial estimate and observation
638 availability on convective-scale data assimilation with an ensemble Kalman filter.
639 Mon. Wea. Rev., 132, 1238–1253.
640 Zhang, S. and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates
641 of error covariance on assimilation in a simple atmospheric model. Tellus, 55A, 126-
642 147.
643 Zhang, S., M. J. Harrison, A. Rosati and A. T. Wittenberg, 2007: System design and
644 evaluation of coupled ensemble data assimilation for global oceanic climate studies.
645 Mon. Wea. Rev., 135, 3541-3564.

646 Zhang, S., Z. Liu, A. Rosati and T. Delworth, 2012: A Study of Enhancive Parameter
647 Correction with Coupled Data Assimilation for Climate Estimation and Prediction
648 Using a Simple Coupled Model. Tellus A, 64, 10963,
649 doi:10.3402/tellusa.v64i0.10963.
650 Zhang, S., 2011a: Impact of Observation-Optimized Model Parameters on Decadal
651 Predictions: Simulation with a Simple Pycnocline Prediction Model. Geophys. Res.
652 Lett., 38, L02702, doi:10.1029/2010GL046133.
653 Zhang, S. 2011b: A study of impacts of coupled model initial shocks and state-parameter
654 optimization on climate predictions using a simple pycnocline prediction model.
655 Journal of Climate, 24(23),, 6210–6226. doi: http://dx.doi.org/10.1175/JCLI-D-10-
656 05003.1

657 **Figure Caption**

658 Fig. 1. The model monthly SST response to 3-m SPD uncertainty at different month
659 for (a) March, (b) June, (c) September and (d) December. The shading represents the
660 correlation coefficient between the SPD ensemble and the 1st month monthly SST
661 response while the contours represent the magnitude of the monthly SST response
662 (ensemble spread). A 30-member ensemble simulation that starts from the same initial
663 condition but use different values of the parameter SPD. The SPD ensemble is
664 constructed as a Gaussian distribution with the mean of 20-m and the standard deviation
665 of 3-m. We integrate the model from the beginning of each month to the end of the month
666 to obtain the monthly mean response.

667 Fig. 2. Solar penetration depths estimated using DAEPC with the GPO method. The
668 total ensemble size is 30. Panel (a) and (c) are the spatial distribution of parameter error
669 values and parameter ensemble spreads after 20 years simulation for the perfect

observation experiment. Panel (b) and (d) are the parameter errors and parameter ensemble spreads after 20 years simulation for regular observation experiments.

Fig. 3. The estimated SPD after 40-year simulations using DAEPC with the GPO method. Panels (a) and (b) are for EXP-1a using perfect observations: (a) the scatter diagram between SPD error values and its ensemble spreads. The red line is the regression line. (b) The blue line is the averaging value of SPD using top percent grids (with smallest ensemble spread) and the red dish lines represent 1-standard deviation of the averaging values. The black dish line is the "truth".
Panels (c) and (d) are the same as (a) and (b) but for EXP-1b using regular observations.

Fig. 4. The Estimated SPD using DAEPC with the ASA (EXP-2a) and SA (EXP-2b). (a)Temporal evolution of parameter error (thick lines) and 1-standard deviation of ensemble spread (thin lines). The red lines are for EXP-2b and the blue lines are for EXP-2a the green lines are also for EXP-2a but with a reduced minimum parameter ensemble spread of 0.2 for the year 31~47. The black solid line is the "truth" and the black dish lines are the minimum parameter ensemble spreads (uncertainty goals) for the experiments. (b) temporal evolution of total numbers of grids used for average in ASA.

Fig. 5. (a) the scatter diagram between SPD error values and its ensemble spreads for EXP-2a after the 1st analysis cycle of parameter updating. The red line is the regression; (b) is the same as (a) but for after the 5th analysis cycle. (c) the numbers of "good" grids (values) for the 1-5 analysis cycles of EXP-2a using ASA with different threshold. The blue line is for the 1^{st} analysis cycle, the green line is for the 2^{nd}, the red line is for 3^{rd}, the cyan line is for 4^{th} and the magenta is for 5^{th}.
(d) the mean SPD values of the "good" grids from (c) respectively.

693 Fig. 6. (a) the temporal evolution of SPD (thick lines) and 1-standard deviation of
694 ensemble spread (thin lines) for EXP-3. The red lines are for EXP-3b and the blue lines
695 are for EXP-3a. The black solid line is the "truth" and the black dish lines are the
696 minimum parameter ensemble spreads (uncertainty goals) for the experiments.
697 (b) is same as (a) but for EXP-4.
698 Fig. 7. (a) the temporal evolution of m_d (thick lines) and 1-standard deviation of
699 ensemble spread (thin lines) for EXP-5. The red lines are for EXP-5b and the blue lines
700 are for EXP-5a. The black solid line is the "truth" and the black dish lines are the
701 minimum parameter ensemble spreads (uncertainty goals) for the experiments.
702 (b) is same as (a) but for the temporal evolution of m_q for EXP-6.
703 Fig. 8. The scale (variance) of each term in equation (A12). The blue cure is for
704 the scale of the 1st ($\sigma_\beta N_\beta^f(1-\theta)$) at the right hand of equation, which is related to the
705 error of parameter prior; the green curve is the scale of the 2nd term ($\sigma_\beta\sqrt{\theta-\theta^2}N_x^b$) at the
706 right hand of equation, which is related to the uncertainties of observation and forecast
707 but unrelated to the parameter uncertainty. The black curve is the scale of the total error
708 (ε_β^a).
709
710
711
712
713
714
715

716

717

718

719

720 Table 1 the experiment setting. The oceanic observations are SST and SSS; and

721 atmospheric observations are T, U and V. EXP-1a uses the perfect observations (truth).

722 EXP-5 a & b estimate the parameter of m_d and EXP-6 a & b estimate the parameter of

723 m_q.

EXP	Method	Obs. (Ocn.; Atm.)	Parameter (Truth)	Initial guess/Truth/ uncertainty
1a &b	GPO	1 month; 1 day	SPD	18m/17m/1m for SPD
2a	ASA	1 month; N/A	N/A	20m/17m/3m for SPD
2b	SA	1 month; N/A	N/A	20m/17m/3m for SPD
3a	ASA	10 days; N/A	N/A	20m/17m/3m for SPD
3b	SA	10 days; N/A	N/A	20m/17m/3m for SPD
4a	ASA	1 day; N/A	N/A	20m/17m/3m for SPD
4b	SA	1 day; N/A	N/A	20m/17m/3m for SPD
5a	ASA	1 month	N/A	1.2/1.0/0.2 for m_d
5b	SA	1 month	N/A	1.2/1.0/0.2 for m_d
6a	ASA	1 month	N/A	1.2/1.0/0.2 for m_q
6b	SA	1 month	N/A	1.2/1.0/0.2 for m_q

724

725

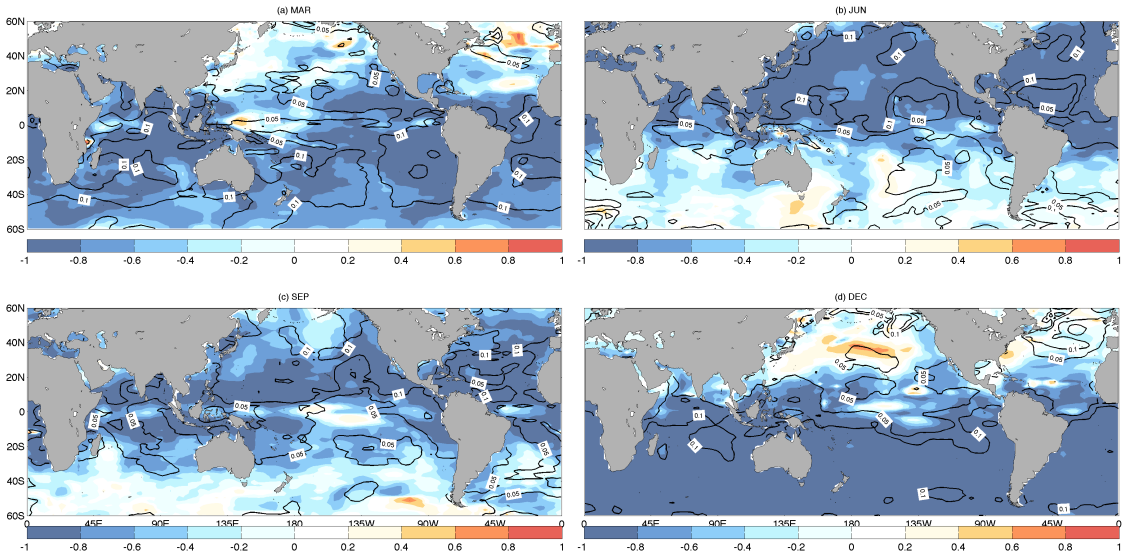


Fig. 1. The model monthly SST response to 3-m SPD uncertainty at different month for (a) March, (b) June, (c) September and (d) December. The shading represents the correlation coefficient between the SPD ensemble and the 1st month monthly SST response while the contours represent the magnitude of the monthly SST response (ensemble spread). A 30-member ensemble simulation that starts from the same initial condition but use different values of the parameter SPD. The SPD ensemble is constructed as a Gaussian distribution with the mean of 20-m and the standard deviation of 3-m. We integrate the model from the beginning of each month to the end of the month to obtain the monthly mean response.

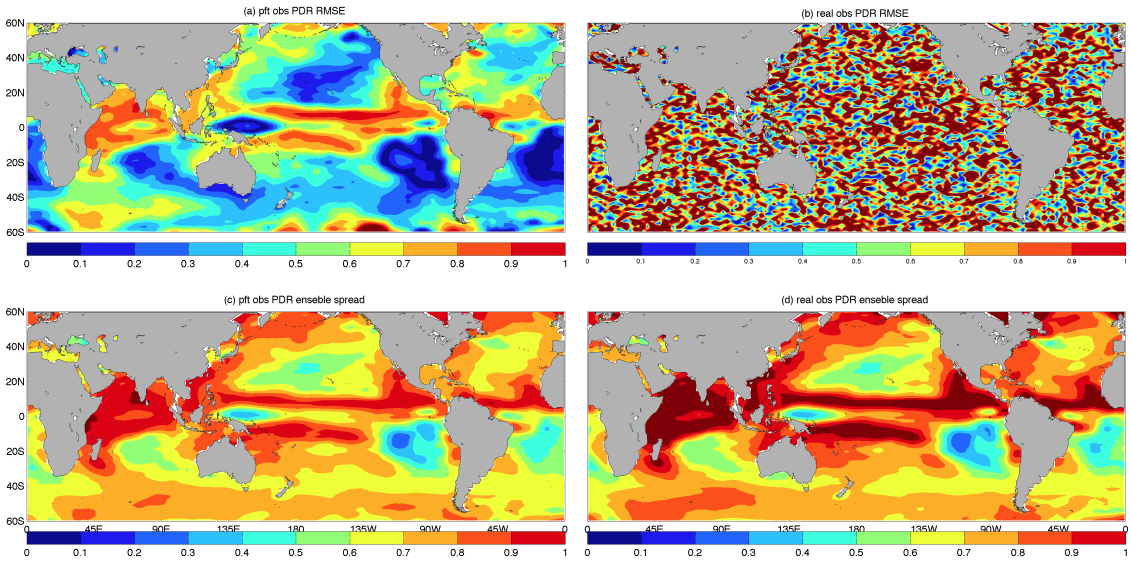


Fig. 2. Solar penetration depths estimated using DAEPC with the GPO method. The total ensemble size is 30. Panel (a) and (c) are the spatial distribution of parameter error values and parameter ensemble spreads after 20 years simulation for the perfect observation experiment. Panel (b) and (d) are the parameter error values and parameter ensemble spreads after 20 years simulation for regular observation experiments.

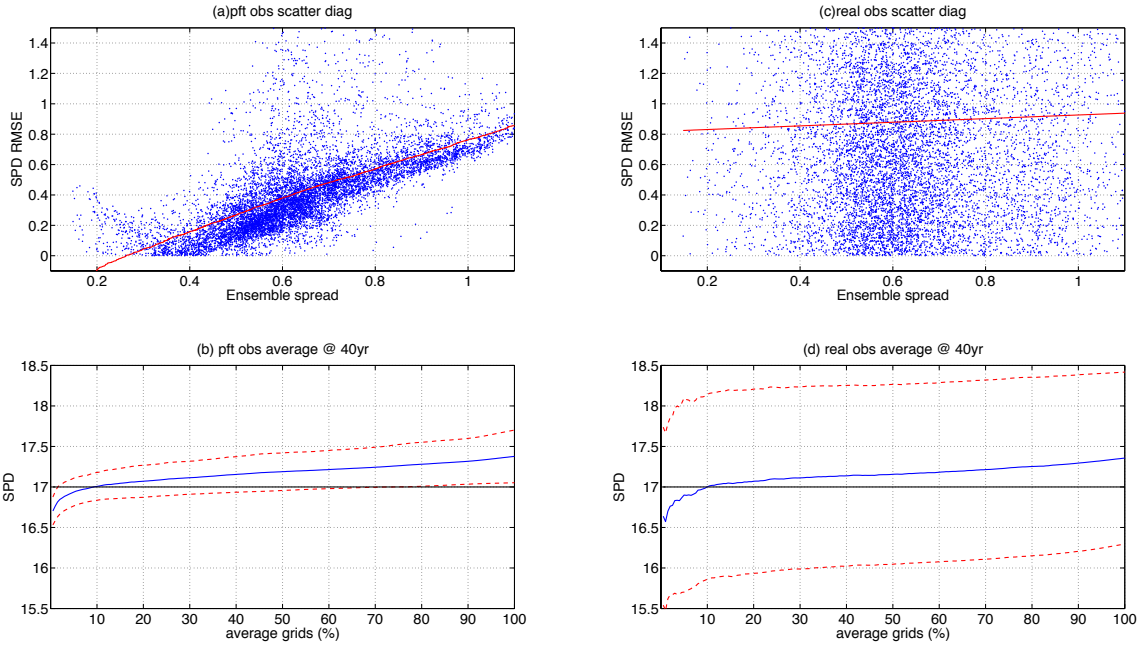
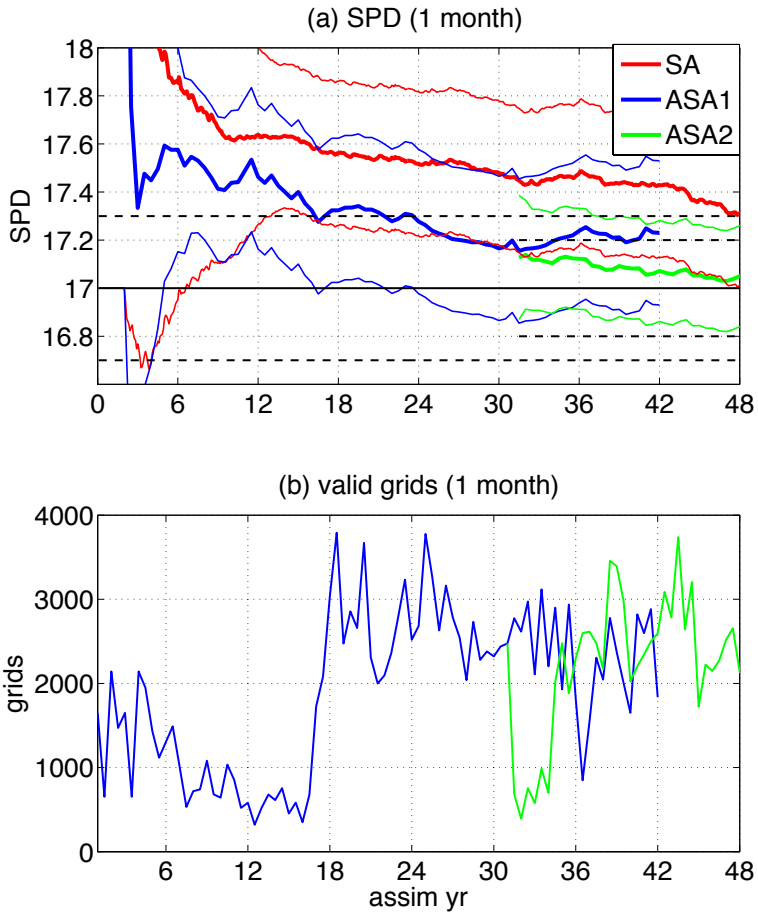


Fig. 3. The estimated SPD after 40-year simulations using DAEPC with the GPO method. Panels (a) and (b) are for EXP-1a using perfect observations: (a) the scatter diagram between SPD error values and ensemble spreads. The red line is the regression line. (b) The blue line is the averaging value of SPD using top percent grids (with smallest ensemble spread) and the red dish lines represent 1-standard deviation of the averaging values. The black dish line is the "truth".

Panels (c) and (d) are the same as (a) and (b) but for EXP-1b using regular observations.



782
783
784

785 Fig. 4. The Estimated SPD using DAEPC with the ASA (EXP-2a) and SA (EXP-2b).

786 (a)Temporal evolution of parameter error (thick lines) and 1-standard deviation of

787 ensemble spread (thin lines). The red lines are for EXP-2b and the blue lines are for EXP-

788 2a the green lines are also for EXP-2a but with a reduced minimum parameter ensemble

789 spread of 0.2 for the year 31~47. The black solid line is the "truth" and the black dish

790 lines are the minimum parameter ensemble spreads (uncertainty goals) for the

791 experiments. (b) temporal evolution of total numbers of grids used for average in ASA.

792

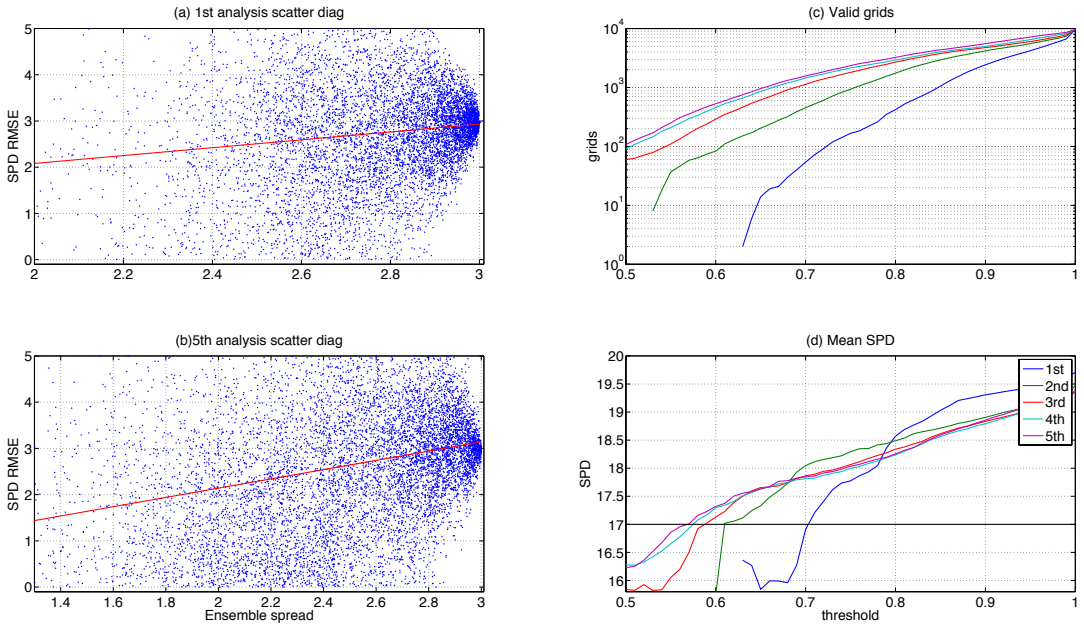


Fig. 5. (a) the scatter diagram between SPD error values and its ensemble spreads for EXP-2a after the 1st analysis cycle of parameter updating. The red line is the regression line. (b) is the same as (a) but for after the 5th analysis cycle.

(c) the numbers of "good" grids (values) for the 1-5 analysis cycles of EXP-2a using ASA with different threshold. The blue line is for the 1st analysis cycle, the green line is for the 2nd, the red line is for 3rd, the cyan line is for 4th and the magenta is for 5th.

(d) the mean SPD values of the "good" grids from (c) respectively.

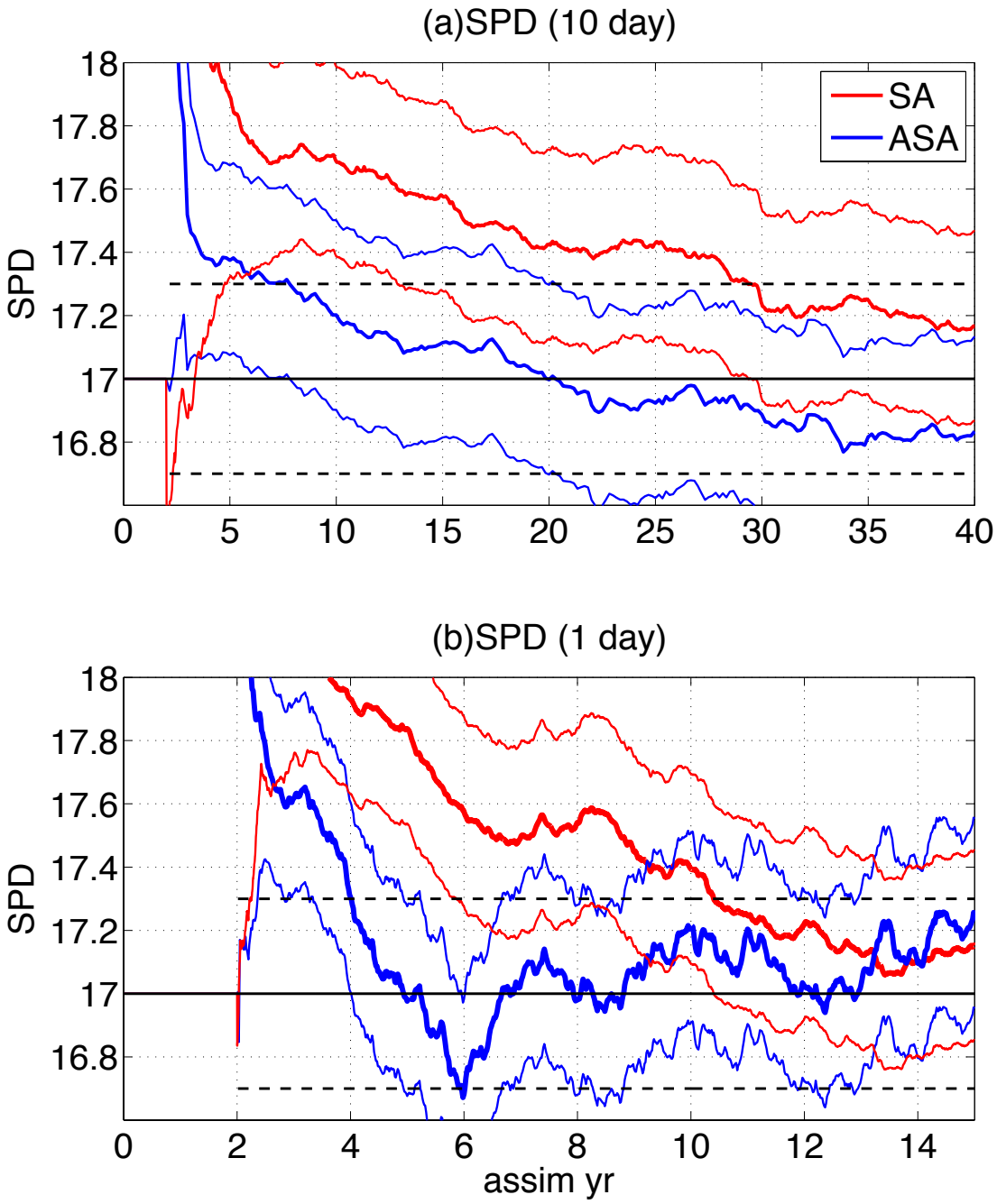
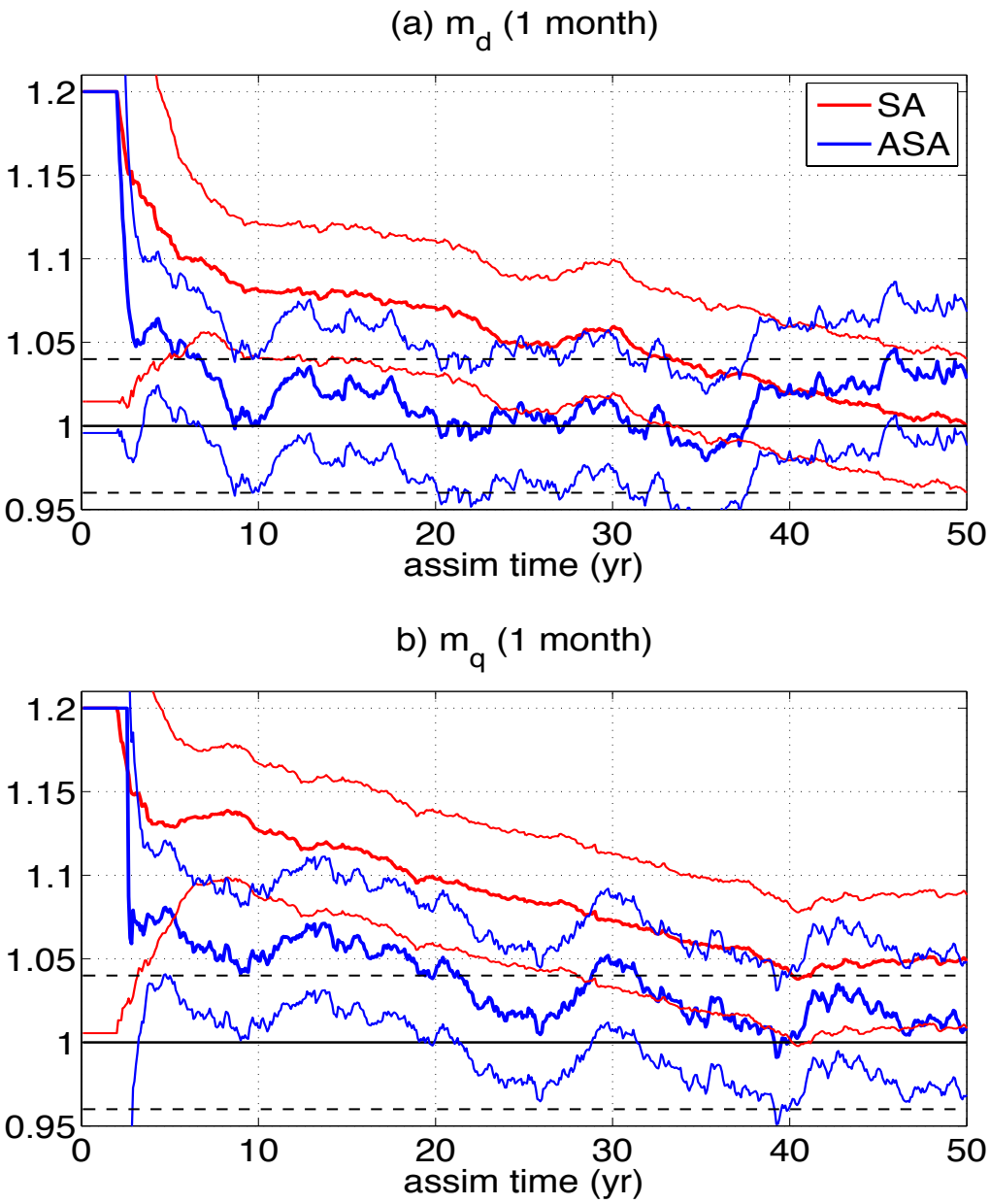


Fig. 6. (a) the temporal evolution of SPD (thick lines) and 1-standard deviation of ensemble spread (thin lines) for EXP-3. The red lines are for EXP-3b and the blue lines are for EXP-3a. The black solid line is the "truth" and the black dish lines are the minimum parameter ensemble spreads (uncertainty goals) for the experiments. (b) is same as (a) but for EXP-4.



814

815 Fig. 7. (a) the temporal evolution of m_d (thick lines) and 1-standard deviation of

816 ensemble spread (thin lines) for EXP-5. The red lines are for EXP-5b and the blue lines

817 are for EXP-5a. The black solid line is the "truth" and the black dish lines are the

818 minimum parameter ensemble spreads (uncertainty goals) for the experiments.

819 (b) is same as (a) but for the temporal evolution of m_q for EXP-6.

820

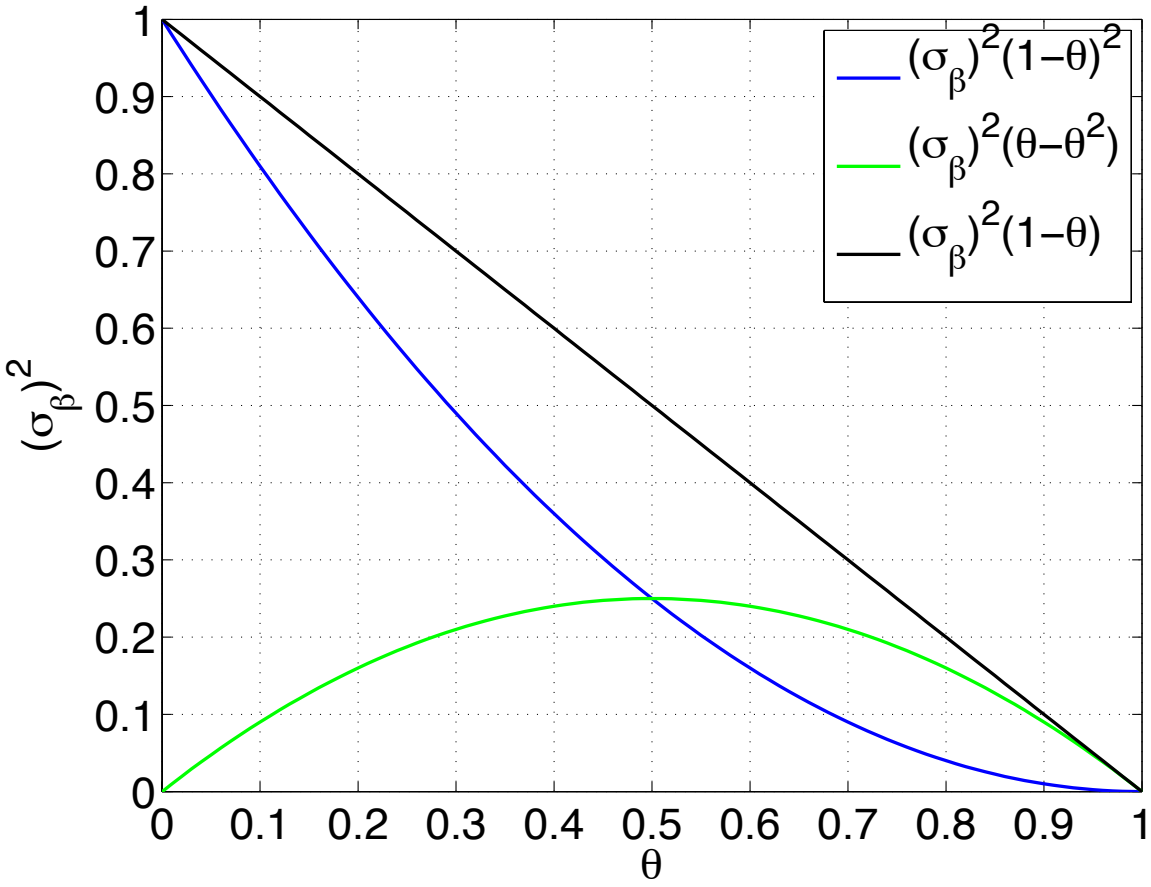


Fig. 8. The scale (variance) of each term in equation (A2). The blue curve represents the scale of the 1st ($\sigma_\beta N_\beta^f(1-\theta)$) at the right hand side of the equation, which is related to the error of the parameter prior; the green curve represents the scale of the 2nd term ($\sigma_\beta\sqrt{\theta-\theta^2}N_x^b$) at the right hand side of equation, which is related to the uncertainties of the observation and forecast, but unrelated to the parameter uncertainty. The black curve is the scale of the total error (ε_β^a).